

INSTITUTE FOR  
MATHEMATICS



STRUCTURES  
CLUSTER OF  
EXCELLENCE



UNIVERSITÄT  
HEIDELBERG  
ZUKUNFT  
SEIT 1386

Michael Bleher

*Institute for Mathematics, Heidelberg University*

COMPUTATIONAL TOPOLOGY - NEURAL NETWORKS MEETING, SEVILLA 2023

# TOPOLOGICAL SIGNATURES OF CONVERGENCE IN VIRAL EVOLUTION

based on

arXiv:2106.07292

arXiv:2207.03394

Joint w/

Andreas Ott, Maximilian Neumann (Karlsruhe)

Lukas Hahn (Heidelberg)

Juan Patiño-Galindo (Mount Sinai)

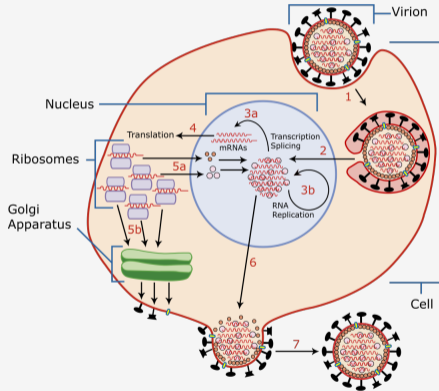
Mathieu Carrière (Inria Sophia-Antopolis)

Raul Rabadan (Columbia)

Ulrich Bauer (Munich)

Samuel Braun, Holger Obermaier, Mehmet Soysal, René Caspart (Karlsruhe)

# A Brief Introduction to Genomics and Epidemiology



Author: YK Times, Wikimedia Commons (CC BY-SA 3.0)

## Viral Genome

Sequence of nucleotides *A, C, T, G*.  
Encodes instructions for host cell.

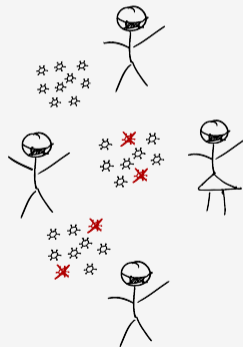
## Viral Life Cycle

1. Virus binds to host cell
2. Viral genome enters cell & nucleus
3. Replication and Transcription of viral RNA
4. Translation (*production of viral proteins*)
5. & 6. Assembly
7. Release

# A Brief Introduction to Genomics and Epidemiology

## Viral Transmission

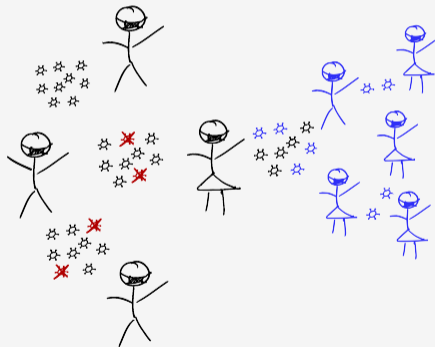
- not every mutation is beneficial



# A Brief Introduction to Genomics and Epidemiology

## Viral Transmission

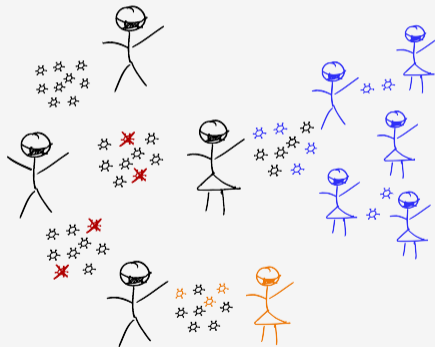
- not every mutation is beneficial
- mutations that spread widely are not necessarily beneficial (founder effects)



# A Brief Introduction to Genomics and Epidemiology

## Viral Transmission

- not every mutation is beneficial
- mutations that spread widely are not necessarily beneficial (founder effects)
- not every beneficial mutation catches on

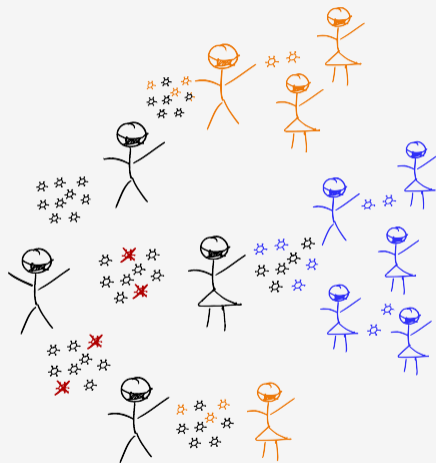


# A Brief Introduction to Genomics and Epidemiology

## Viral Transmission

- not every mutation is beneficial
- mutations that spread widely are not necessarily beneficial (founder effects)
- not every beneficial mutation catches on
- BUT: beneficial mutations tend to *appear repeatedly* (and may then spread more widely)

**Recurrence is a hallmark of adaptation**



# Topology of Viral Evolution

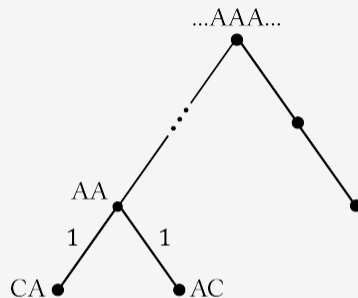
Monitor evolution of virus and determine influence of (single or groups of) mutations.

Construct **phylogenetic tree**.

## Basic Idea:

Hamming distance = Tree distance

Minimum spanning tree reconstructs ancestral relations





# Topology of Viral Evolution

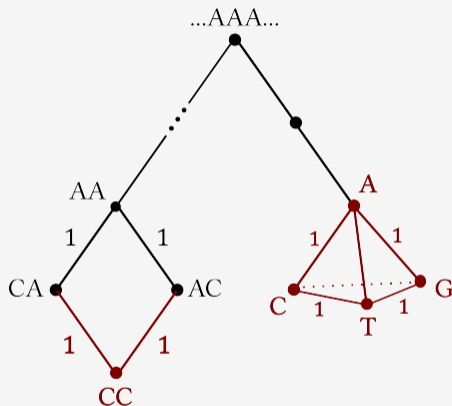
Monitor evolution of virus and determine influence of (single or groups of) mutations.

Construct **phylogenetic network**.

## Basic Idea:

Hamming distance  $\neq$  Tree distance

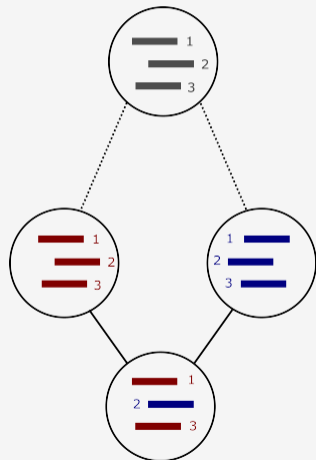
Minimum spanning tree reconstructs ancestral relations, **but is not unique**.



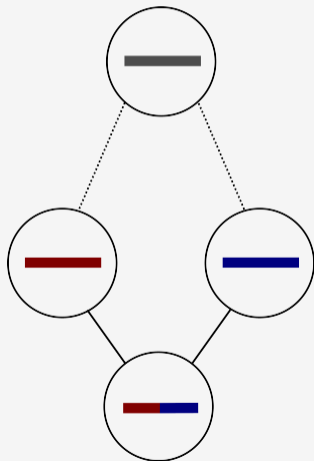
# Topology of Viral Evolution

## Reassortment

Some viruses have disconnected genome, e.g. Flu (HxNy).  
Co-infection can lead to "reassortment" during assembly.



# Topology of Viral Evolution



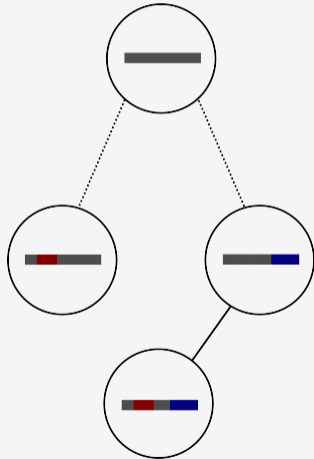
## Reassortment

Some viruses have disconnected genome, e.g. Flu (HxNy).  
Co-infection can lead to "reassortment" during assembly.

## Recombination

Replication apparatus can switch template.  
Co-infection can lead to recombination into a hybrid genome.

# Topology of Viral Evolution



## Reassortment

Some viruses have disconnected genome, e.g. Flu (HxNy).  
Co-infection can lead to "reassortment" during assembly.

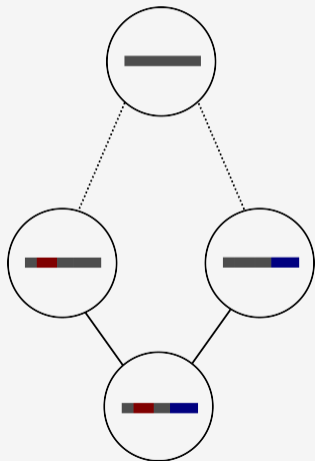
## Recombination

Replication apparatus can switch template.  
Co-infection can lead to recombination into a hybrid genome.

## Convergence / Homoplasy

independent emergence of similar features.  
example: evolution of flight (mammals, insects, bats)

# Topology of Viral Evolution



## Reassortment

Some viruses have disconnected genome, e.g. Flu (HxNy).  
Co-infection can lead to "reassortment" during assembly.

## Recombination

Replication apparatus can switch template.  
Co-infection can lead to recombination into a hybrid genome.

## Convergence / Homoplasy

independent emergence of similar features.  
example: evolution of flight (mammals, insects, bats)

⇒ **cycles in phylogenetic network at different scales.**

# Persistent Homology

Consider genomic data with Hamming distance as finite metric space  $(X, d)$ .

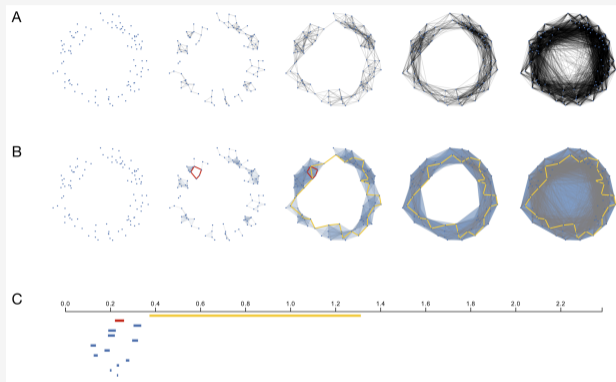
Construct Vietoris-Rips complex

$$VR_{\bullet}(X, d)$$

Cycles at scale  $r$  correspond to 1d homology  $H_1(VR_r(X, d))$ .

Persistent Homology  $\simeq$  homology at all scales simultaneously  $H_k(VR_{\bullet}(X, d))$ .

*NB: Calculation of homology is just matrix reduction. Highly optimized, e.g. Ripser (Uli Bauer).*



# Persistent Homology

## Contractibility Lemma(s)

### Rips, Gromov (60's & 80's)

$(X, d)$  a  $\delta$ -hyperbolic geodesic metric space  $\implies \text{VR}_t(X)$  is contractible,  $t \geq 4\delta$ .

### Chan, Carlsson, Rabadan (2013)

If  $(X, d)$  is a tree, then  $H_n(\text{VR}_\bullet(X, d)) = 0, n \geq 1$ .

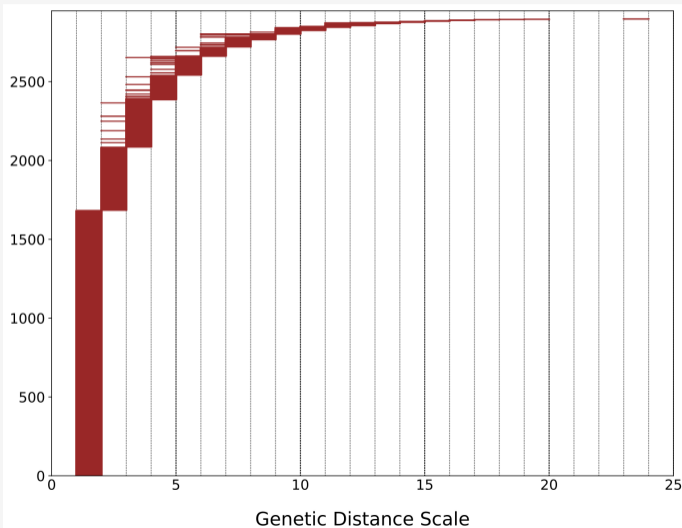
### Bauer, Roll (2022)

$(X, d)$  a  $\delta$ -hyperbolic  $\nu$ -geodesic finite metric space  $\implies \exists$  discrete gradient collapse:

$$\text{VR}_s(X) \searrow \text{VR}_t(X) \searrow \{*\}, \quad s > t \geq 4\delta + 2\nu$$

$\implies$  **Persistent homology detects evolutionarily relevant phenomena!**

# Persistent Homology of SARS-CoV-2



*February 28th 2021*

~ 450,000 isolates

$|H_1| \sim 2,600$



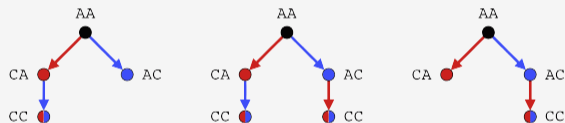
# Noise or Feature?

## Back-of-the-envelope

$$p \simeq 2/30,000 \simeq \mathcal{O}(10^{-4})$$

$$\# \text{unique sequences} = \mathcal{O}(10^6)$$

→ expect  $\mathcal{O}(10^2)$  cycles due to noise.



# Noise or Feature?

## Back-of-the-envelope

$$p \simeq 2/30,000 \simeq \mathcal{O}(10^{-4})$$

$$\# \text{unique sequences} = \mathcal{O}(10^6)$$

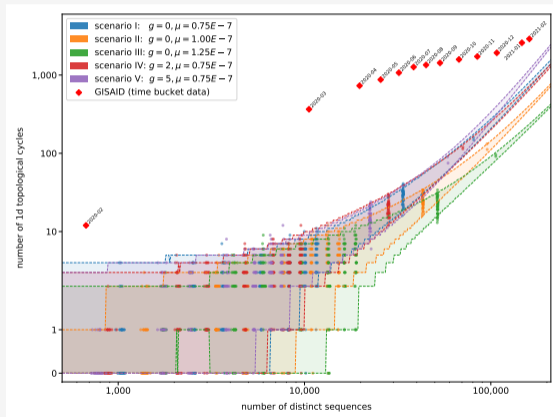
→ expect  $\mathcal{O}(10^2)$  cycles due to noise.

## Simulations of neutral evolution

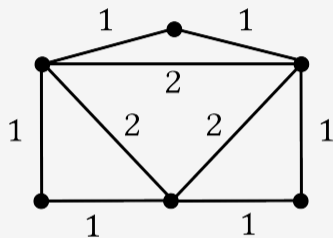
- uniform mutation probability
- No fitness advantage
- No recombination

→ expect 350-400

(worst case:  $\leq 1,200$ )



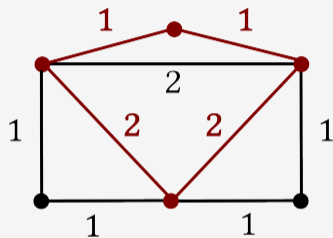
# The topological Recurrence Index (tRI)



Which mutations give rise to homology?

use **exhaustive** cycle representatives

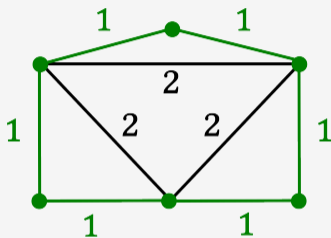
# The topological Recurrence Index (tRI)



Which mutations give rise to homology?

use **exhaustive** cycle representatives

# The topological Recurrence Index (tRI)



Which mutations give rise to homology?

use **exhaustive** cycle representatives

Exhaustive representatives of  $[1, x)$ -persistent classes have only edges of length 1: every edge corresponds to a single nucleotide variation (SNV).

$[1, x)$ -persistent classes = "SNV-cycles".

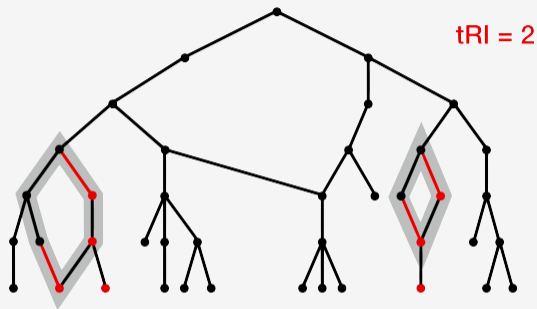
# The topological Recurrence Index (tRI)

$\{\gamma\}$  – exhaustive representatives of SNV-cycles

$\mu$  – mutation (xNy, e.g. D614G)

$$\text{tRI}(\mu) := \#\{\gamma \mid \mu \in \gamma\}$$

(without double counting edges)

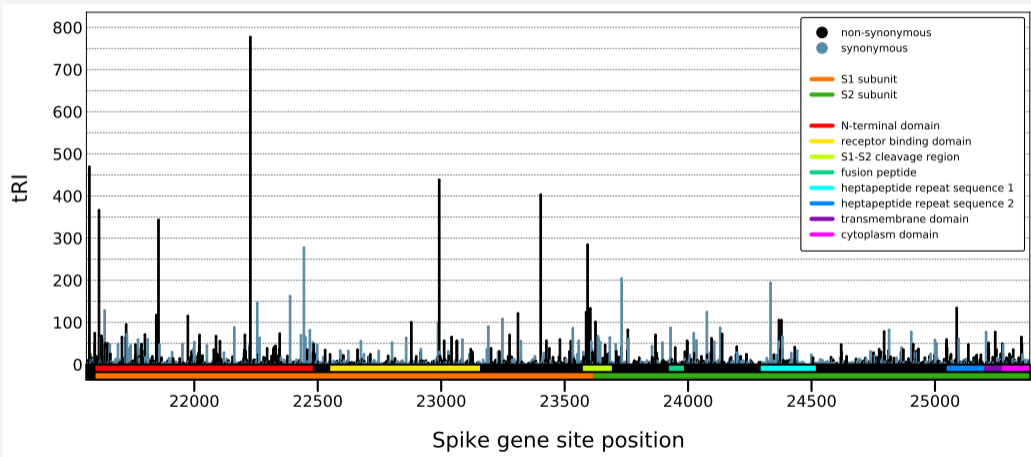


## Proposition

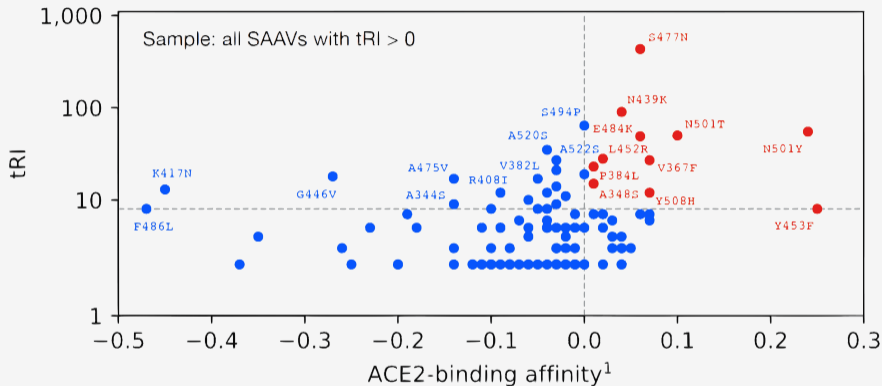
$\text{tRI}(\mu)$  = minimal number of independent occurrences of  $\mu$  in  $X$ .

$\implies$  **tRI is a measure for convergence.**

# Topological Recurrence of Spike mutations



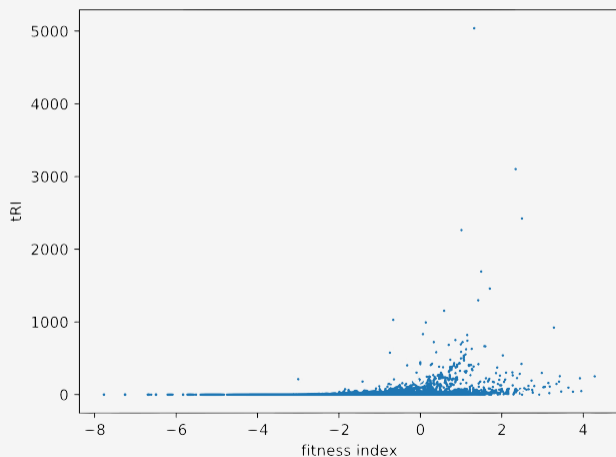
# Correlation with Host Adaptation



**significant tRI ( $\geq 8$ ) correlates with increase in binding affinity.**



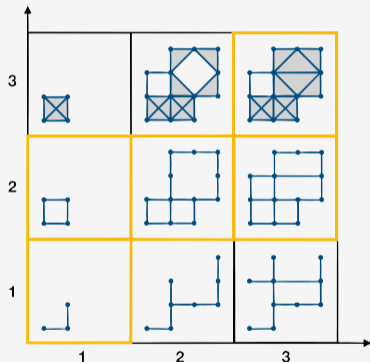
# Comparison with Fitness Index



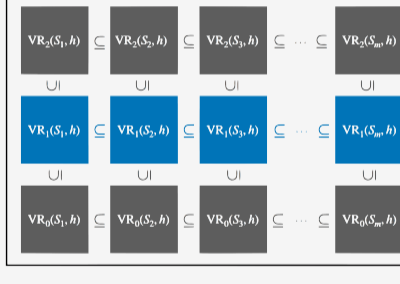
**positive tRI correlates with positive fitness index (recently introduced by Bloom & Neher).**

# A Word on Multipersistance and a Computational Trick

Time series data, investigate 2d-persistence module.

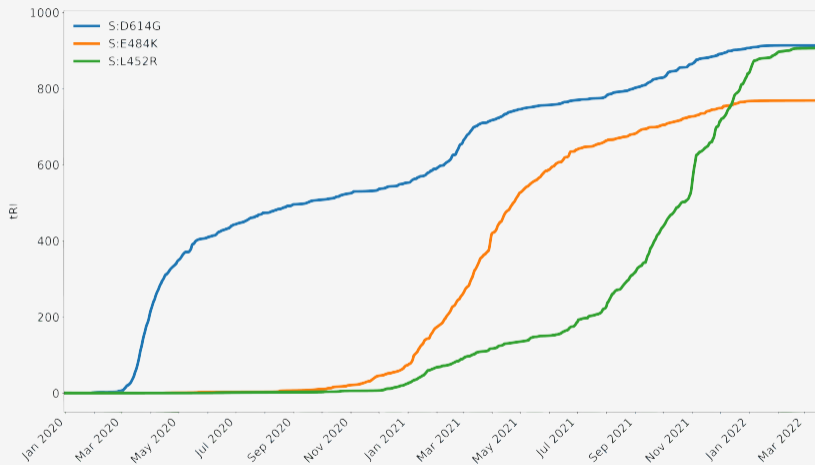


Vietoris-Rips scaling  
parameter

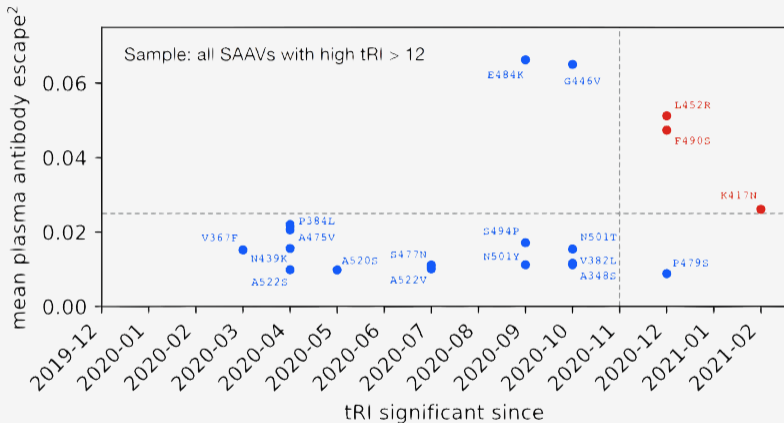


Trick: Restriction to 1d submodule is equivalent to deformation of metric (generally leads to semi-metric, deformation violates triangle inequality)

# tRI Curves



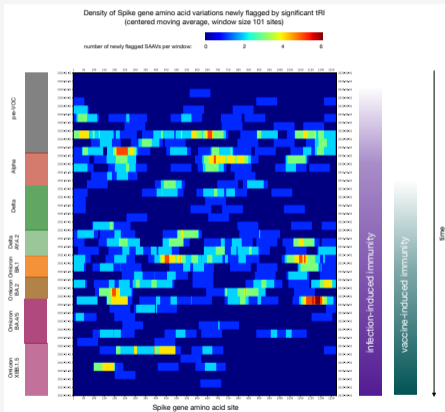
# Evolutionary Dynamics and Epistasis



*Acquisition date of significant tRI correlates with immune escape.*

-> late 2020 evolutionary driving force shifts from transmission to immune escape

# Evolutionary Dynamics and Epistasis



smoothed tRI growth rate along the genome shows surprising amount of time-dependence.

Looks like measure of epistasis: influence of given mutational background on fitness of newly acquired mutations.

# Summary

- Persistent homology detects evolutionarily relevant phenomena
- topological Recurrence Index (tRI) detects adaptive mutations (among others)
- tRI computations are efficient
- tRI curves might allow study of epistasis
- Differentiation of beneficial and adversarial mutations must rely on experiments, but persistent homology can tell us where to look

# Summary

- Persistent homology detects evolutionarily relevant phenomena
- topological Recurrence Index (tRI) detects adaptive mutations (among others)
- tRI computations are efficient
- tRI curves might allow study of epistasis
- Differentiation of beneficial and adversarial mutations must rely on experiments, but persistent homology can tell us where to look

Thank you!